VSB — TECHNICAL UNIVERSITY OF OSTRAVA

FACULTY OF ECONOMICS

DIPLOMA THESIS

2019 Siyu Liu

VSB — TECHNICAL UNIVERSITY OF OSTRAVA

FACULTY OF ECONOMICS

DEPARTMENT OF FINANCE

Property Prices: Analyzing the Effect of Macroeconomic Determinants

Ceny nemovitostí: Analýza vlivu makroekonomických determinant

Student: Bc. Siyu Liu

Supervisor of the diploma thesis: doc. Ing. Aleš Melecký, Ph.D

Ostrava 2019

VŠB - Technical University of Ostrava
Faculty of Economics
Department of Finance

Diploma Thesis Assignment

Student: **Bc. Siyu Liu**

Study Programme: N6202 Economic Policy and Administration

Study Branch: 6202T010 Finance

Title: Property Prices: Analyzing the Effect of Macroeconomic Determinants
Ceny nemovitostí: Analýza vlivu makroekonomických determinant

The thesis language: English

Description:

1. Introduction
2. Property Prices and Their Determinants
3. Methodology of Regression Analysis
4. Empirical Analysis of Macroeconomic Determinants of Property Prices
5. Conclusion

Bibliography
List of Abbreviations
Declaration of Utilisation of Results from the Diploma Thesis
List of Annexes
Annexes

References:

GUJARATI, Damodar N. *Basic Econometrics.* 4th ed. New York: McGraw-Hill, 2003. 1002 p. ISBN 9780072335422.
BROOKS, Chris and Sotiris TSOLACOS. *Real Estate Modelling and Forecasting.* 1st ed. Cambridge: Cambridge University Press, 2010. 474 p. ISBN 9780521873390.
DE BANDT, Olivier and Thomas KNETSCH. *Housing Markets in Europe: A Macroeconomic Perspective.* 1st ed. Berlin: Springer, 2010. 406 p. ISBN 9783642153396.

Extent and terms of a thesis are specified in directions for its elaboration that are opened to the public on the web sites of the faculty.

Supervisor: **doc. Ing. Aleš Melecký, Ph.D.**

Date of issue: 23.11.2018
Date of submission: 26.04.2019

Ing. Iveta Ratmanová, Ph.D.
Head of Department

prof. Dr. Ing. Zdeněk Zmeškal
Dean

The declaration

"I hereby declare that I have elaborated the entire thesis including annexes myself."

Ostrava date 22.4.2019

Siyu Liu 刘思雨

Student's name and surname

Content

1. Introduction

Property is a significant part of the entire social wealth, and fluctuations of property prices affect the stability and development of the financial system and macro economy. As we know, property purchases are one of the highest expenditures for individuals and households. The fluctuations of property prices are closely connected with daily lives of people. For corporations, real estates are important parts of corporate assets. In the different level of enterprises, many of them pledge real estates to obtain loans. Property industry is involved with energy, materials, construction and many other important industries of the national economy.

The goal of this thesis is to analyze the effect of macroeconomic determinants on property prices. We study this topic because property prices are among the interests of central banks especially after the global financial crisis, when financial stability became separate policy in most developed central banks. This topic is also relevant for investors and households in many countries.

This thesis aims to determine the influence of selected variables on property prices for 15 European countries during the period from 2008 to 2018. It not only theoretically discusses the impact of property market supply and demand factors on property prices, but also empirically studies how various economic variables determine real estate prices through panel data models.

The structure of this thesis is as follows. There are five chapters in total. The first one is the introduction. It introduces our topic and objective of this thesis, and gives the brief description of individual chapters. In Chapter 2, we describe property market and macro factors that affect property price. In this part, we introduce our dependent variable and independent variables for our analysis later. Chapter 3 is a theoretical part. We introduce the definition of panel data and methodology on how we can proceed panel data regression analysis. Chapter 4 discusses the empirical analysis of selected variables introduced in Chapter 2 on sample of selected European countries. We use methods of panel data analysis to estimate our models and choose the best one to discuss the economic interpretation. The last chapter is conclusion. It briefly describes the procedure of this thesis and summarize the findings and provide conclusions arising from the thesis with regard to the objective specified in the introduction.

2. Property Prices and Their Determinants

As we know, property industry (or residential property industry) influences various aspects of national economy, it affects individuals, corporations and nations.

The research on housing market and housing prices has always been a significant part of economics research. In this chapter, we would briefly discuss the residential property market and most of macroeconomic determinants of property price.

Main source of this chapter is from *De Bandt et al. (2010).*

2.1 Property market

According to *Oxford English Dictionary (2011)*, real estate or residential property is "*property consisting of land and the buildings on it, along with its natural resources such as crops, minerals or water; immovable property of this nature; an interest vested in this (also) an item of real property, (more generally) buildings or housing in general*."

Property market refers to all exchanges about residential properties. This market consists of supply side and demand side, which is similar to other goods and services markets. Compared of other markets, there are still differences on this market, further we would discuss the specifics.

2.1.1 Participants on property market

There are plenty of participants on this market. They are mainly real estate owners, tenants, property developers, second-hand transferers and related institutions.

On demand side, usages for real estate owners can be subdivided into three types: self-occupation, commercial use and investment. For self-occupation and commercial use, owners take advantage of original functions of real estate. Owners can either reside in this house or use the house for commercial purposes, such as opening a restaurant. For investment, properties are regarded as investment products. Investors can rent them to others or wait until properties' value rises to sell them for making profits. Tenants usually are low-income people or young adults who have not worked too long. They cannot afford to buy property because of the high payments of mortgages.

As for supply side, property developers are major parties on property market. They provide different houses (usually they are new) depending on different needs of customers. In contrast, although second-hand house transferers are also suppliers on this market, they only transfer the residential properties that belongs to them to someone else. This will not inject new properties into the market, nor increase the total amount of properties.

Institutions can refer to either demand party or supply party. They can also link the supply and demand sides of property to provide services to all participants on the entire market.

Institutions related to this market include banks, real estate agents, law firms, etc. On the one hand, banks and other financial institutions provide mortgages credit for those who lack money to buy properties on demand side. On the other hand, when buyers are unable to repay the loan, banks will take the houses and sell their properties on the market, i.e. banks also affect supply side. Real estate agents help to match the supply and demand sides and receive commissions from the process. Law firms can help buyers and sellers to develop contracts to reduce friction in transactions.

2.1.2 Characteristics of property market

In general, there are six basic characteristics on property and property market.

First, house is a durable product that can usually last for decades or even hundreds of years when no accidents happen. The supply on this market is relatively fixed in the short-run. At any time, the supply of real estates mainly depends on the current real estate stock, deterioration rate of existing real estates and newly developed real estates. Therefore the size of real estate stock cannot be adjusted in short periods, and the impact of this market adjustments is often weakened by large-scale stock.

Second, residential properties are heterogeneous. The geographical environment, traffic conditions, architectural style and other characteristics of the houses are unique, and even individual rooms in the same building have big differences. In addition, properties with different functions are also very diverse. For example, shopping malls in prosperous areas, high-end office buildings and residential communities with good environment are completely different residential properties. Therefore, the substitutability of properties is very low.

Third, transaction costs of buying and moving to a new house is much higher than most transaction costs of the same type. This transaction costs includes brokerage, search costs, legal and administrative costs, statutory cost, financing costs and so on.

Fourth, the market has a long time lag. There are certain lags in all aspects of the property market, such as financing channels, housing design and construction, especially in supply side. The production cycle of property goods is very long, generally several years. The longer production cycle determines the lag of property supply, which in turn leads to the risk of supply. Even if the development plan of property is feasible at present, when the house is put into the market a few years later, it may also cause backlogs and slow sales due to changes in the market. Compared with the general market, the adjustment mechanism of the real estate market is slower. On the other hand, in demand side, there are also lags between between decisive factor changes and house price changes. Most individuals and families do not have the foresight to buy a house before the price rises. They can only respond if they actually feel the impact of these factors on their lives.

Fifth, property is both commodity and an investment. It can be used in daily life or be regarded as an investment for profit. This dual nature is likely to cause the consumer's investment exceeding the actual value of property, which will lead to excessive investment in the market.

Sixth, the location of existing properties cannot be moved. The space is fixed means that market adjustments must be moved by consumers, and consumers need to choose where they want from existing properties. For example, demand preferences of people changes. They may work in downtown but prefer to live in the suburb. Therefore, they will search for houses in the suburb with good connection to city center. However, the existing properties in the city center cannot be moved to the suburb.

2.2 Determinants of property price

In general, determinants of property price can be divided into two aspects: micro aspects and macro aspects.

In micro aspects, most of researches such as *Lamont et al. (1999), Genesove et al. (2001), Case et al. (1988)* and so on, which start from the perspective of customers. They consider the impact of a range of factors that influence customers' purchase decisions on property prices. As for macro aspects, researchers pay more attention to the impact of macro indicators on prices and analyze the relationship between property market and macro economy. For instance, *Follain et al. (1982, p. 13)* studies the relationship between inflation and property price; *Englund et al. (1997, p. 17)* shows the relationship between GDP growth and property price. In this thesis, we focus on macroeconomic determinants of property prices.

2.2.1 Supply and demand determinants

When it comes to supply of a product, the first thing we need to consider is its cost. Cost is one of the most important factors affecting the market. The costs of property generally consists of the following components: price of lands, administrative expenses, costs of construction, costs of materials, interests expenses of loans, taxes and other expenses. All parties on supply side have to ponder the impact of costs. For example, property developers bear most of above costs, and they must estimate their affordable costs when they develop properties. When price of selected lands or other costs are higher, their supply of properties are usually adjusted. Second-hand house transfers mainly bear administrative expenses and transaction costs. These expenses will affect supply as well.

Then, expectation is also important for supply of property. Expectation is mainly about what suppliers expect for property market in the future. It includes all expectations consuming land market, inflation, interest rate, consumers’ needs and so on. For instance, if property developers expect that price of land in a certain area where they plan to build would increase in the future, they are more likely to choose purchase the land now.

Of course, most of the decisive factors, such as costs and expectations, affect housing prices under the influence of both supply and demand.

For demand side, the demographic factor is the most vital factor affecting housing demand. The establishment of new families, population growth, and expansion of family size will lead to a change in demand for residential properties. Especially in China, houses are crucial for newly-married people. From an emotional point of view, most Chinese people hold an engrained traditional belief that they would have a stable life in their own house. When newly-married people can afford a house, it shows the certain social status and economic strength. They can prove to their parents that they are able to take care of themselves.

Also, wages and salaries are significant factors for consumers. As a rule, income is positively correlated with residential demand. When people receive higher wages, they are more likely to choose to live in a better place. For instance, instead of renting a house, people can buy their own house, or people are able to move to a bigger house. With the demand increasing, prices of residential property rise. However, due to the large differences in economic levels of different regions, income has a different impact on property prices.

Other factors, such as the level of economic development and the economic cycle, will affect housing prices from both supply and demand.

2.3.2 Other determining factors

Among the other determining factors, the national policy factors may play important role. Generally, the impact of national policies can be divided into two types——direct and indirect. The direct impacts of national policy indicate that the direct goal of national policy is to regulate real estate prices. For example, in order to curb the rise in real estate prices, China has introduced a more stringent bank lending policy and increased the down payment ratio of second homes, trying to restrain investment-oriented consumers, thereby controlling the irrational rise in house prices.

The indirect impacts of national policies refer to national policies that do not target the real estate market but affect prices of real estate. For instance, when central banks conduct a series of monetary policies to control the money supply and interest rates, they do not mainly aim at property market, but these polices do influence this market.

From a macro perspective, money supply affect property prices. When money supply increases, the amount of local currency increases in money market and domestic interest decreases. Local currency will depreciate under the situation of foreign currency unchanged due to interest rate differential decreasing. On the one hand, the prices of imported materials and products will be higher. These materials and products will lead to price level increases in domestic country. On the other hand, after depreciation of domestic currency properties will be relatively cheaper for foreigners. The situation of foreign capital investment will lead to an increase in demand for real estate. The both effects cause the house prices to rise.

In recent years, the European Central Bank has implemented a loose monetary policy and maintained a very low interest rate, which has supported the rise in housing prices.

Here is a table about mortgage interest rates in EU from the second quarter of 2015 to the third quarter of 2018.

From this table, we can see mainly declining trend of mortgage rates in various 16 European countries. But mortgage rates of the Czech Republic, Hungry, Poland and Romania show different trends. Poland has shown a flat trend during selected period. Other three illustrate mainly rising trend of mortgage rates. Denmark owns the lowest mortgage rate which is 0.79% at the third quarter of 2018 and mortgage rate of Finland is the closest to it with 0.88%.

Table 2.1 Mortgage interest rates (%)

	II 2015	III 2015	IV 2015	I 2016	II 2016	III 2016	IV 2016	I 2017	II 2017	III 2017	IV 2017	I 2018	II 2018	III 2018
BE	2.43	2.46	2.48	2.27	2.05	2.00	2.00	2.11	2.16	2.13	2.03	2.01	2.01	1.95
CZ*	2.48	2.48	2.42	2.37	2.25	2.17	2.01	2.17	2.22	2.17	2.25	2.41	2.49	2.58
DE	1.83	2.03	2.00	1.91	1.81	1.68	1.63	1.80	1.83	1.85	1.83	1.85	1.90	1.87
DK**	1.06	1.03	1.12	1.17	1.29	1.20	1.09	1.11	1.09	0.94	0.87	0.87	0.84	0.79
ES	2.25	2.17	2.08	2.02	2.04	2.02	1.97	1.97	1.92	1.99	1.91	1.95	1.93	1.95
FI	1.51	1.43	1.32	1.23	1.20	1.16	1.16	1.13	1.07	1.02	0.95	0.92	0.87	0.88
FR***	2.01	2.14	2.18	2.02	1.69	1.46	1.32	1.45	1.54	1.55	1.52	1.48	1.45	1.43
HU	5.04	4.57	4.85	4.88	4.58	4.34	4.06	3.91	3.59	3.43	3.01	4.31	4.31	4.87
IE	3.46	3.40	3.42	3.30	3.34	3.26	3.22	3.16	3.22	3.20	3.07	3.02	3.06	2.97
IT	2.77	2.67	2.50	2.33	2.20	2.02	2.02	2.11	2.10	2.02	1.90	1.88	1.80	1.80
NL	2.92	2.90	2.83	2.75	2.64	2.54	2.39	2.40	2.40	2.40	2.40	2.30	2.30	2.30
PL	4.30	4.40	4.40	4.40	4.50	4.40	4.40	4.40	4.40	4.40	4.40	4.30	4.30	4.40
PT	2.28	2.19	2.13	1.99	1.86	1.76	1.77	1.70	1.61	1.48	1.52	1.51	1.41	1.33
RO****	3.95	3.94	3.79	3.49	3.32	3.56	3.52	3.71	3.34	3.60	4.42	4.78	5.01	5.67
SE	1.55	1.52	1.56	1.62	1.60	1.59	1.57	1.65	1.52	1.53	1.56	1.52	1.51	1.48
UK	2.60	2.57	2.54	2.50	2.41	2.30	2.16	2.09	2.05	1.98	1.98	2.03	2.09	2.10

* For Czech Republic from Q1 2015 the data source is the Czech national Bank.

** This data series has been revised and it depicts the variable interest rate, which is the most common one.

*** Data from Q2 2012 has been revised for France due to a new source. Further data break in Q1 2014.

**** Recalculation of the interest rate as a weighted average of interest rates in local currency and euro. (previously weighted average only of euro denominated mortgages). Data break from Q1 2014.

Source: European Mortgage Federation.

Available on https://hypo.org/emf/press-release/emf-publishes-quarterly-review-q3-2018/

2.2.3 Selected variables

There are one dependent variable and six independent variables in our model. Our aim is to analyze the relationship among them.

2.2.3.1 Dependent variable

The dependent variable is house prices index. According to *Eurostat*, house price index reflects price changes of all residential properties purchased by households. The data are expressed as quarterly index (2015=100).

2.2.3.2 Independent variables

We choose below independent variables.

CPI (Consumer Price Index): CPI is an index that captures the change in the prices of a basket of consumer goods and services in a country. It is used to measure inflation, which is the crucial index for price level. When general price level goes up, property price increases as well.

GDP(Gross Domestic Product): GDP has always been an important indicator of a country's economic development, and also it shows the economic cycle of a country. In general, the increase in GDP stands for a country's increased wealth, which stimulate consumers' demand for real estates. If suppliers will not increase the supply of real estate efficiently without extra costs. The price of houses in this country will rise. This industry is an important part of GDP growth.

Population: it measures the demographic factor which affects demand for properties in selected country. It is a significant factor on demand side of property market. In general, demand of property increases as the population increases.

Unemployment rate: unemployment affects property prices mainly from two ways. On the one hand, unemployed people have no salaries. In a country, if unemployment rate is too high, demand of property is insufficient, which causes housing prices go down. On the other hand, excessive unemployment rate has a depressing effect on real estate investment, while low unemployment rate will stimulate real estate investment, which in turn affects housing prices.

Income: it is used to measure wages and salaries of people. As we mentioned before, it may be closely correlated with demand for properties. The income level directly measures the purchasing power of consumers.

Interest rate: it refers to the price of money. On supply side, when interest rate rises, loan rate is higher and the cost of property developers to obtain funds becomes higher, and their profit margin may be lower. During this situation, developers may reduce the development of properties or increase property price. The former is more likely to cause a decline in supply of property; the latter directly increase the price. On demand side, when interest rate rises, the increase of interest on bonds and bank deposits will attract residents to invest bonds and deposits. Further, higher interest rates transmit to lending rates and make bonds more expensive, thus reduce demand for properties. In other words, demand of consumers on property market may decline and property price will probably decrease.

In daily life, most homebuyers purchase house through mortgage. The mortgage interest rate will directly affect the expenditures which people plan to buy the house.

3. Methodology of Regression Analysis

In this chapter, we will introduce the methodology of regression analysis which we will use in the next chapter.

When we want to analyze the effect of macroeconomic determinants on property prices, we need to study the relationship between macroeconomic determinants and property prices at different time points in various European countries. We choose to use panel data analysis here and try to form a function to specify the casual relationship between house prices and variables which we have mentioned at the end of last chapter.

Main sources of this chapter are from *Maddala (2001)* and *Wooldridge (2002)*.

3.1 Panel data

3.1.1 Definition

In econometrics, panel data is multi-dimensional data involving measurements over time. Panel data contains observations of multiple phenomena obtained over multiple time periods for the same individuals.

Time series and cross-sectional data can be considered as special cases of panel data which are in one dimension only. Time series data are collected on one individual over several time periods. Cross-sectional data are collected on several individuals at one point in time.

For instance, in this thesis, the data of a country's housing prices index and determinants over time are time series data, while data from different countries at a certain time point are cross-sectional data. This panel data includes N countries which are observed at T regular time periods. In our model, panel data are long panel which means we have more time periods than countries ($N < T$).

We make a simplified example below to show how panel data looks like. There are only three countries (A,B,C) and four years. All countries are observed in all time periods. (This data set is called balanced panel.) Y represents dependent variable which is house price index in this thesis. X_1, X_2, and X_3 refer to independent variables, and in our model we have more variables than here.

Table 3.1 Panel data example

N (Country)	T (Time)	Y	X_1	X_2	X_3
A	2015				
A	2016				
A	2017				
A	2018				
B	2015				
B	2016				
B	2017				
B	2018				
C	2015				
C	2016				
C	2017				
C	2018				

Source: own elaboration.

3.1.2 Stationarity of data

Stationary time series are more likely to help us to figure out a permanent, rather than temporary, rule between the dependent and independent variables.

There are two types of stationary time series. First, strict stationarity. It requires that all moments of the joint distribution of stochastic variables process are invariant to time shifts. The condition is too strict to be applied in economy.

Second, weak stationarity. Time series should fulfill following conditions:

1. The mean of the process is constant and equal to a specific number;

2. The variance of the process is time invariant and equal to finite constant;

3. The covariance of the process should not be time dependent, it can be affected just by the distance between the two time stick considered.

In this sub-chapter, we would discuss several ways to test the stationarity of data and if the data is non-stationary, how can we transform the time series.

3.1.2.1 Unit root tests

The unit root test is a special method for testing the stationarity of data. When the autoregressive lag polynomial has one root α equal to one, we say it has a unit root.

$$y_t = \alpha y_{t-1} + \varepsilon_t, |\alpha| < 1. \tag{3.1}$$

Where y_t is variables which we need to test, α is the root, and ε_t is the error term.

There are many methods for unit root test, including ADF test, PP test and NP test. Here, we focus on ADF test.

An augmented Dickey-Fuller test (ADF) tests the null hypothesis that a unit root is present in a time series sample which means $\alpha = 1$. The alternative hypothesis is different depending on which version of the test is used, but is usually stationarity or trend-stationarity. We prefer to accept alternative hypothesis.

First of all, we describe Dickey-Fuller (DF) test briefly. Based on Formula (3.1), there are three main versions of the test:

1. Testing for a unit root (there is no constant and no trend). We subtract y_{t-1} from both sides of the model and rewrite this model as

$$\Delta y_t = y_t - y_{t-1} = (\alpha - 1) y_{t-1} + \varepsilon_t = \delta y_{t-1} + \varepsilon_t. \tag{3.2}$$

Where Δy_t is the first difference of y_t, and δ is equal to α -1.

2. Testing for a unit root with drift (there is constant but no trend).

$$\Delta y_t = c_0 + \delta y_{t-1} + \varepsilon_t. \tag{3.3}$$

Where c_0 is the drift.

3. Test for a unit root with drift and deterministic time trend (there are both constant and trend).

$$\Delta y_t = c_0 + c_1 t + \delta y_{t-1} + \varepsilon_t. \tag{3.4}$$

Where $c_1 t$ is the deterministic time trend.

These models can be used to estimate and test for a unit root is equivalent to testing δ = 0.

ADF test adds lagged differences to the model to eliminate autocorrelation of residuals. We would discuss autocorrelation later. Here, the testing procedure for the ADF test is applied to the model.

$$\Delta y_t = c_0 + c_1 t + \delta y_{t-1} + \theta_1 \Delta y_{t-1} + \ldots + \theta_{j-1} \Delta y_{t-j+1} + \varepsilon_t. \quad (3.5)$$

Where j is the lag order of the autoregressive process, and θ is the coefficient of difference of y_t.

And ADF statistic used in the test, is a negative number. The more negative it is, the stronger the rejection of the hypothesis that there is a unit root at some level of confidence.

3.1.2.2 KPSS test

There are other ways testing stationarity of time series except unit root tests. For instance, KPSS test.

The Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test figures out if a time series is stationary around a mean or linear trend, or is non-stationary due to a unit root. The null hypothesis for the test is that the data is stationary. The alternate hypothesis is that the data is not stationary.

Contrary to most unit root tests, the presence of a unit root is not the null hypothesis but the alternative. In addition, in the KPSS test, the absence of a unit root is not a proof of stationarity but trend-stationarity. This is an important distinction since it is possible for a time series to be non-stationary, have no unit root but be trend-stationary.

The KPSS test is based on linear regression. It breaks up a series into three parts: a deterministic trend ($c_1 t$), a random walk (r_t), and a stationary error (ε_t), with the regression equation

$$y_t = r_t + c_1 t + \varepsilon_t. \quad (3.6)$$

If the data is stationary, it will have a fixed element for an intercept, or the series will be stationary around a fixed level.

3.1.2.3 Transformation of time series

When time series are non-stationary, we may always use some techniques to transform them to be stationary. Here are three simple techniques.

First, we can difference the data. In other words, we can use the change in data between two periods. That is, based on the given series Z_t, we create the new series. For instance, first difference can be calculated as

$$Y_t = Z_t - Z_{t-1}. \tag{3.7}$$

The differenced data will contain one less observation than the original data.

Second, if the series contains a trend, we can fit some type of curve to the data and then calculate the residuals as a difference between actual data and fitted curves.

Third, for non-constant variance, taking the logarithm or square root of the series may stabilize the variance. For negative data, we can add a suitable constant to make all the data positive before applying the transformation. This constant can then be subtracted from the model to obtain predicted (i.e., the fitted) values and forecasts for future points.

3.2 Panel data models

In general, there are two types of panel data models: the pooled regression model, and the individual-specific effects model.

3.2.1 Pooled regression model

The key assumption of pooled regression model is that there are no unique attributes of individuals within the measurement set, and no universal effects across time. This regression model specifies constant coefficients, referring to both intercepts and slopes.

$$y_{it} = c + \mathbf{x}_{it}\beta + u_{it}. \tag{3.8}$$

Where y_{it} is the dependent variables observed for individual i at time t, c is the unobserved constant coefficient, $\mathbf{x}_{it}$ is the time-variant $T \times k$ (the number of independent variables) regressor matrix, β is the $k \times 1$ matrix of parameters, and u_{it} is the error term.

If there is no significant difference between different individuals in terms of time; for the cross-section, there is no significant difference between different sections, then we can directly use pooled ordinary least squares(OLS) to analyze this model.

3.2.2 Individual-specific effects models

The individual-specific effects are the leftover variation in the dependent variables that cannot be explained by the regressors.

We assume that there is unobserved heterogeneity across individuals which are generated by c_i (i means different individuals). For example, unobserved ability of a person that affects wages.

There are two common assumptions which are made about the time-invariant individual effects c_i. They are the random effects assumption and the fixed effects assumption. The random effects assumption (made in a random effects model) is that the individual-specific effects are uncorrelated with the independent variables. The fixed effects assumption (made in a fixed effects model) is that the individual-specific effects are correlated with the independent variables.

The fixed effects (FE) model allows the individual-specific effects to be correlated with the regressors $\mathbf{x}$. We include c_i as intercepts. Hence, each individual has a different intercept term and the same slope parameters.

$$y_{it} = c_i + \mathbf{x}_{it}\beta + u_{it}. \tag{3.9}$$

Since c_i is not observable, it cannot be directly controlled for. We can get the individual effects after estimation as:

$$\hat{c}_i = \bar{y}_i - \bar{\mathbf{x}}_i\hat{\beta}. \tag{3.10}$$

Where $\hat{c}_i$ is one of estimates of c_i from (3.9), $\bar{y}_i$ is the mean of the dependent variables observed for individual *i*, $\bar{\mathbf{x}}_i$ is the mean of time-invariant regressor matrix, and $\hat{\beta}$ is an estimate of β.

In the random effects (RE) model, the individual effects are randomly distributed across the cross-sectional units and in order to capture the individual effects, the regression model is specified with an intercept term representing an overall constant term. RE analysis puts c_i into the error term. We can state assumptions

$$E(u_{it} \mid \mathbf{x}_{it}, c_i) = 0, t = 1, \ldots, T. \tag{3.11}$$

$$E(c_i \vert \mathbf{x}_{it}) = E(c_i) = 0. \tag{3.12}$$

In many applications the whole point of using panel data is to allow for c_i to be arbitrarily correlated with the $\mathbf{x}_{it}$. A fixed effects analysis achieves this purpose explicitly.

In this thesis, we focus more on pooled regression model and fixed effects regression model.

3.3 Ordinary least squares

The panel data model can be estimated with several estimators. We believe our estimators are consistent and efficient. Consistency is established based on the law of large numbers. If an estimator is consistent, more observations will tend to provide more precise and accurate estimates. Efficiency means we want estimators have the lowest possible variance.

OLS is a type of linear least squares method for estimating the unknown parameters in a linear regression model. The pooled OLS estimator is generated by stacking the data over i and t into one long regression with $N \times T$ observations and estimating it by OLS.

$$y_{it} = \beta_0 + \beta_1 \mathbf{x}_{it} + \beta_2 \mathbf{x}_{it} + \ldots + \beta_k \mathbf{x}_{it} + u_{it}. \tag{3.13}$$

Where β_k is the $k \times 1$ matrix of parameters, and k is the number of independent variables.

This model appears overly restrictive because β is the same in each time period. The two assumptions sufficient for pooled OLS to consistently estimate β are as follows

1. $E(\mathbf{x}_t' u_t) = 0, t = 1, 2, \ldots, T.$ (3.14)
2. $rank\left[\sum_{t=1}^{T} E(\mathbf{x}_t' \mathbf{x}_t)\right] = K.$ (3.15)

Where E means expected value, $\mathbf{x}_t^{'}$ is the transpose matrix of $\mathbf{x}_t$, T is the total time periods, and K is the total number of independent variables.

From above two assumptions, there is nothing about the relationship between $\mathbf{x}_s$ and u_t for $s \neq t$, and there is perfect linear dependencies among the explanatory variables.

3.3.1 Statistical verification of the parameters and model

In this part, we will explain how to do statistical verification of β_k and verification of the whole model. We use t-test to test hypotheses about individual regression slop coefficient. Tests of more than one coefficient at a time are typically done with the F-test. And in following two tests we prefer to accept alternative hypothesis.

The t-test is appropriate to use when the stochastic error term is distributed normally and when he variance of that distribution must be estimate.

Here is hypotheses of t-test for β_k.

H_0: $\beta_k = 0$ (β_k - coefficient is not statistically significant).

H_1: $\beta_k \neq 0$ (β_k - coefficient is statistically significant).

For the null hypothesis to be rejected, an observed result has to be statistically significant, i.e. the observed p-value is lower than the pre-specified significance level. In other words, calculated t-value is higher than critical one which we figure out in STATA. Here is how we define the t-statistic.

$$t_{calculate} = \frac{\hat{\beta}_k - \beta_k}{\sqrt{s^2 (\mathbf{x}_t^{'} \mathbf{x}_t)^{-1}}}. \tag{3.16}$$

Where $\hat{\beta}_k$ is an estimate of β_k and s^2 is the variance of sample.

T-test can be used only for testing one hypothesis at a time. Further, if we want to add more coefficients (or restrictions) in the hypotheses, we need to use F-test. If the restrictions are true (null hypothesis), then the restricted model fits the data in the same way as the unrestricted model residuals are nearly the same. If the restrictions are false, then the restricted model fits the data poorly residuals from the restricted model are much higher than those from the unrestricted model. The basic idea is to compare the sum of squared residuals from the two models and find out if their difference is positive or equal to zero. If the difference is positive, then we reject null hypothesis that the restrictions are true.

H_0: $\beta_1 = \beta_2 = ... = \beta_k = 0$ (the whole model is not statistically significant).

H_1: $\beta_1 \neq 0 \vee \beta_2 \neq 0 \vee ... \vee \beta_k \neq 0$ (the whole model is statistically significant).

The F-statistics for testing Q linear restrictions on the $K\times 1$ vector $\boldsymbol{\beta}$ is

$$F = \frac{(SSR_r - SSR_{ur})}{SSR_{ur}} \cdot \frac{(NT-K)}{Q}. \tag{3.17}$$

Where SSR_r is the sum of squared residuals from regression using the NT observations with the restrictions imposed, SSR_{ur} is the sum of squared residuals from the full unrestricted model, and Q is the number of restrictions.

3.3.2 Autocorrelation

Autocorrelation is correlation between members of series of observations ordered in time (as in time series data) or space (as in cross-sectional data). It is the similarity between observations as a function of the time lag between them.

It is often useful to have a simple way to detect autocorrelation after estimation by pooled OLS. One reason to test for autocorrelation is that it should not be present if the model is supposed to be dynamically complete in the conditional mean. A second reason is to see whether we should compute a robust variance matrix estimator for the pooled OLS estimator. We need to detect if there is autocorrelation in our model. If autocorrelation exits we have to find out how to eliminate it.

We focus on the alternative that the error is a first-order autoregressive process; this will have power against fairly general kinds of autocorrelation.

The function of residuals written as

$$u_t = \rho_1 u_{t-1} + e_t. \tag{3.18}$$

Where

$$E(e_t | \mathbf{x}_t, u_{t-1}, \mathbf{x}_{t-1}, u_{t-2}, ...) = 0. \tag{3.19}$$

Under the null hypothesis of no autocorrelation, $\rho = 0$. The alternative hypothesis is: $\rho \neq 0$ (there is autocorrelation).

One way to proceed is to write the dynamic model under autoregressive model (3.16) as:

$$y_t = \mathbf{x}_t \beta + \rho_1 u_{t-1} + e_t, t = 2,...,T. \tag{3.20}$$

Where we lose the first time period due to the presence of u_{t-1}. If we can observe the u_t, we know how to proceed: simply estimate Equation (3.18) by pooled OLS (losing the first time period) and perform a t-test on $\hat{\rho}_1$. To carry out this procedure, we replace the u_t with the pooled OLS residuals. Therefore, we run the regression y_{it} on $x_{it}, \hat{u}_{i,t-1}$ $(t = 2,...,T; i = 1,...,N)$. And we operate a standard t-test on the coefficient of $\hat{u}_{i,t-1}$.

In case we find out autocorrelation in the model, there are several approaches to remove autocorrelation from the original (3.11):

1. To include lagged explained variable and use h-statistic.

$$y_{it} = \beta_0 + \beta_1 \mathbf{x}_{it} + \beta_2 \mathbf{x}_{it} + ... + \beta_k \mathbf{x}_{it} + \beta_{k+1} y_{i,t\text{-}1} + u_{it}. \tag{3.21}$$

2. To include lagged explanatory variables.

$$y_{it} = \beta_0 + \beta_1 \mathbf{x}_{i,t-1} + \beta_2 \mathbf{x}_{i,t-2} + ... + \beta_k \mathbf{x}_{i,t-k} + u_{it}. \tag{3.22}$$

3. To include trend variable if it is in residuals.

$$y_{it} = \beta_0 + \beta_1 \mathbf{x}_{it} + \beta_2 \mathbf{x}_{it} + ... + \beta_k \mathbf{x}_{it} + \beta_{k+1} trend_{it} + u_{it}. \tag{3.23}$$

4. To use the Cochrane-Orcutt iterative procedure.
5. To estimate using generalized least squares (GLS) methods, New West estimation.

3.3.2 Heteroscedasticity

A collection of random variables is heteroscedastic means there are sub-populations that have significantly different variabilities from others. Heteroscedasticity is the absence of homoscedasticity (the assumption of homoscedasticity refers to that the errors term u_t in regression equation have a common variance σ^2). The existence of heteroscedasticity is a major concern in the application of regression analysis, including the analysis of variance, as it can invalidate statistical tests of significance that assume that the modelling errors are uncorrelated and uniform—hence that their variances do not vary with the effects being modeled.

The null hypothesis of homoscedasticity can be expressed as $E(u_t^2|\mathbf{x}_t)=\sigma^2,\ t=1,2,...,T$. Under this hypothesis, u_{it}^2 is uncorrelated with any function of $\mathbf{x}_{it}$. We can use Breusch–Pagan test to detect heteroscedasticity of the model.

Suppose that variance of residuals $V(u_t) = \sigma^2$. If there are some variables $z_1,z_2,...,z_k$, that influence the error variance and if $\sigma_t^2 = f(\lambda_0 + \lambda_1 z_{1t} + \lambda_2 z_{2t} ... + \lambda_k z_{kt})$, then the hypothesis of this test are:

H_0: $\lambda_1 = \lambda_2 = ... = \lambda_k$ (homoscedasticity).

H_1: H_0 is not true (heteroscedasticity).

We would like to accept the null hypothesis.

It is a chi-squared test: the test statistic is distributed $n\chi^2$ with k degrees of freedom. If the test statistic has a p-value below an appropriate threshold (e.g. $p<0.05$) then the null hypothesis of homoscedasticity is rejected and heteroscedasticity assumed.

If the Breusch–Pagan test shows that there is conditional heteroscedasticity, one could either use weighted least squares (if the source of heteroscedasticity is known) or use heteroscedasticity-consistent standard errors.

3.3.3 Multicollinearity

Multicollinearity refers to a situation where a number of explanatory variables in a multiple regression model are closely correlated with one another. When the explanatory variables are highly intercorrelated, it becomes difficult to disentangle the separate effects of each of the explanatory variables on explained variable. It can lead to skewed or misleading results when a researcher or analyst is trying to estimate how well each one of a number of individual independent variables can most effectively be utilized to predict or understand the dependent variable in a statistical model.

Multicollinearity does not lower the predictive power or reliability of the entire model, at least in the sample dataset; it only affects the calculation of individual predictors. In other words, a multivariate regression model with collinear predictors can indicate the extent to which the entire predictor bundle predicts outcome variables, but it may not give valid results for any single predictor, or about which predictors are redundant with respect to others.

There are two basic ways to estimate multicollinearity in the regression model. One of them is to check correlation matrix. When all values of correlation between independent variables are lower than 0.8, which means no multicollinearity is in this model. We prefer to use this way in Chapter 4.

The other way is called variance inflation factor (VIF). The VIF is the ratio of variance in a model with multiple terms, divided by the variance of a model with one term alone. The VIF is expressed as:

$$VIF(\hat{\beta}_j) = \frac{1}{1 - R_j^{\,2}}. \tag{3.24}$$

Where: R_t^2 is the squared multiple correlation coefficient for the regression of matrix $\mathbf{x}_j$ on the other covariates (a regression that does not involve the explained variable).

This ratio reflects all other factors that influence the uncertainty in the coefficient estimates. The VIF equals 1 ($R_j^2 = 0$) when the vector $\mathbf{x}_j$ is orthogonal to each column of the design matrix for the regression of $\mathbf{x}_j$ on the other covariates. By contrast, the VIF is greater than 1 when the vector $\mathbf{x}_j$ is not orthogonal to all columns of the design matrix for the regression of $\mathbf{x}_j$ on the other covariates. The VIF is higher than 10, means there is strong multicollinearity in the model.

3.3.4 Normality of residuals

Residuals can be regarded as elements of variation unexplained by the fitted model. Since this is a form of error, the same general assumptions apply to the group of residuals that we typically use for errors in general: one expects them to be roughly normal and independently distributed with a mean of 0 and some constant variance.

The normal distribution of residuals should be symmetric, and has a bell-shape with a peakedness and tail-thickenss leading to a kurtosis of 3. Thus we can test for departures form normality by checking the skewness and kurtosis from a sample data. If skewness is not close to zero , and kurtosis is not close to 3 we would reject the normality of the population.

Since we are concerned about the normality of the error terms, we can create a normal probability plot of the residuals,such as histogram plot or or quantile-quantile plot, If the resulting plot is approximately linear, we proceed assuming that the error terms are normally distributed.

Then, we can carry on non-parametric tests, such as $\chi 2$ test, Jarque-Bera test, Kolmogorov-Smirnov test, and so on.

In this thesis, we choose Jarque–Bera (JB) test to detect whether sample data have the skewness and kurtosis matching a normal distribution. If the data comes from a normal distribution, the JB statistic asymptotically has a chi-squared distribution with two degrees of freedom, so the statistic can be used to test the hypothesis that the data are from a normal distribution.

The hypotheses are below.

H_0 : residuals are normal distributed.

H_1 : residual are not normal distributed.

For validity H_0, we test

$$JB = n*[\frac{s^2}{6} + \frac{(k-3)^2}{24}] \sim \chi^2(2). \tag{3.25}$$

Where n is number of observation, s is skewness and k is kurtosis.

Decision rule is : $JB > \chi^2_\alpha(df) \Rightarrow$ we reject the null hypothesis, i.e. residuals are not normal distributed at a certain significant level.

3.4 Fixed effects regression model

In this part, we will discuss steps and methods in testing the fixed effects model.

3.4.1 Consistency of the fixed effects estimator

Consider the fixed effects model for T time periods in Equation (3.9), we can rewrite this equation as

$$y_{it} = c_i \mathbf{j}_T + \mathbf{x}_{it}\beta + u_{it}. \tag{3.26}$$

Where: $\mathbf{j}_T$ is the $T\times 1$ vector of ones. This equation represents a single random draw from the cross section.

The first FE assumption is strict exogeneity of the explanatory variables conditional on c_i. This assumption is identical to RE Assumption (3.11). The key difference is that we do not hold Assumption (3.12). In other words, for fixed effects analysis, $E(c_i \mid \mathbf{x}_i)$ is allowed to be any function of $\mathbf{x}_i$.

The idea for estimating β under Assumption (3.11) is to transform the equations to eliminate the unobserved effect c_i. To accomplish this purpose, we discuss the fixed effects transformation here (also called the within transformation).

The FE transformation is obtained by first averaging equation (3.9) over $t=1,2,...,T$ to get the cross section equation

$$\bar{y}_i = \bar{\mathbf{x}}_i\beta + c_i + \bar{u}_i. \tag{3.27}$$

Where $\bar{y}_i = T^{-1}\sum_{t=1}^{T} y_{it}$, $\bar{\mathbf{x}}_i = T^{-1}\sum_{t=1}^{T} \mathbf{x}_{it}$, and $\bar{u}_i = T^{-1}\sum_{t=1}^{T} u_{it}$. Subtracting equation (3.27) from Equation (3.9) for each t gives the FE transformed equation:

$$\begin{aligned} y_{it} - \bar{y}_i &= (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)\beta + u_{it} - \bar{u}_i, \text{ or} \\ \ddot{y}_{it} &= \ddot{\mathbf{x}}_{it}\beta + \ddot{u}_{it}, t = 1,2,...,T. \end{aligned} \tag{3.28}$$

This time demeaning of the original equation has removed the individual specific effect c_i.

The FE estimator, denoted by $\hat{\beta}_{FE}$, is the pooled OLS estimator from the regression $\ddot{y}_{it}$ on $\ddot{\mathbf{x}}_{it}$, $t= 1,2,...,T;\ i=1, 2, ...\ N$. We can write equation (3.28) for all time periods as

$$\ddot{\mathbf{y}}_i = \ddot{\mathbf{X}}_i\beta + \ddot{\mathbf{u}}_i. \tag{3.29}$$

Where $\ddot{\mathbf{y}}_i$ is $T\times 1$, $\ddot{\mathbf{X}}_i$ is $T\times K$, and $\ddot{\mathbf{u}}_i$ is $T\times 1$. This set of equations can be obtained by premultiplying equation (3.26) by a time-demeaning matrix.

In order to ensure that the FE estimator is well behaved asymptotically, we need a standard rank condition on the matrix of time-demeaned explanatory variables.

$$rank(\textstyle\sum_{t=1}^{T} E(\ddot{\mathbf{x}}_{it}'\ddot{\mathbf{x}}_{it})) = rank[E(\ddot{\mathbf{X}}_i'\ddot{\mathbf{X}}_i)] = K. \tag{3.30}$$

If $\mathbf{x}_{it}$ contains an element that does not vary over time for any i, then the corresponding element in $\ddot{\mathbf{x}}_{it}$ is identically zero for all t and any draw from the cross section. Because $\ddot{\mathbf{X}}_i$

would contain a column of zeros for all i, Assumption (3.29) could not be true. This assumption shows explicitly why time-constant variables are not allowed in fixed effects analysis.

The fixed effects estimator is also called the within estimator because it uses the time variation within each cross section. On the other hand, the between estimator uses only variation between the cross section observations. It uses the time averages of all variables.

3.4.2 Hausman test

On the basis of above sections, the key reason in choosing between a random effects and fixed effects model is whether c_i and $\mathbf{x}_{it}$ are correlated, it is important to have a method for testing this assumption. *Hausman (1978)* proposed a test based on the difference between the random effects and fixed effects estimates. Since fixed effect is consistent when c_i and $\mathbf{x}_{it}$ are correlated, but random effect is inconsistent, a statistically significant difference is interpreted as evidence against the random effects assumption (3.12). This test evaluates the consistency of an estimator when compared to an alternative, less efficient estimator which is already known to be consistent.

Consider the linear regression model

$$y=b\mathbf{x}+e \tag{3.31}$$

Where y is the dependent variable, $\mathbf{x}$ is vector of regressors, b is a vector of coefficients and e is the error term.

In order to use the OLS procedure, we specify that $\mathbf{x}$ is independent of e. Let H_0 denote the null hypothesis that there is $\mathbf{x}$ and e are independent. The alternative assumption is they are not independent.

To implement Hausman's test, we have to construct two estimators b_0 and b_1, which have the following properties.

b_0 is consistent and efficient under H_0 but is not consistent under H_1;

b_1 is consistent under both H_0 and H_1 but is not efficient under H_0.

Then we consider the difference $q=b_1 - b_0$. Hausman statistic is

$$H = \chi^2[K-1] = q'(Var(b_0) - Var(b_1))^g q. \tag{3.32}$$

Where q' is the transpose matrix of q and g denotes the generalized inverse or the Moore-Penrose pseudoinverse. Under the null hypothesis, this statistic has asymptotically the the chi-squared distribution with the number of degrees of freedom equal to the rank of matrix (Var (b_0) - Var(b_1)).

In panel data, b_1 denotes the vector of random effects estimates without the coefficients on time-constant variables or aggregate time variables, and b_0 denotes the corresponding fixed effects estimates.The difference of RE estimator and FE estimator under two hypotheses is that random effects is preferred under the null hypothesis due to higher efficiency, while under the alternative hypothesis fixed effects is preferred.

Table 3.1 The difference of RE estimator and FE estimator

	H_0 is true.	H_1 is true.
b_1 (RE estimator)	Consistent, efficient	Inconsistent
b_0 (FE estimator)	Consistent, inefficient	Consistent

Source: own elaboration.

3.4.3 Explanatory power of regression model

Explanatory power is the ability of a hypothesis to effectively explain the subject matter it involves in. The determination coefficient R^2 measures the amount of variability in the dependent variable y that can be explained by the explanatory variable.

R-squared is a relative measure and takes values ranging from 0 to 1. When R-squared equals zero, it means our regression line explains none of the variability of the data. When R-squared equals 1, it means our model explains the entire variability of the data.

4. Empirical Analysis of Macroeconomic Determinants of Property Prices

We choose 15 European countries to proceed our analysis. They are Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain, Sweden, United Kingdom. They are old members of European Union before the accession of candidate countries on 1 May 2004.

Data of *CPI* and *house price index* are from *OECD*. Other data come from *Eurostat.* All these data are in quarterly frequency. Time period is from the fourth quarter of 2008 to the third quarter of 2018. There are 40 periods in the sample. In total, number of observations is 600. Details are shown in Table 4.1.

Table 4.1 Descriptive data

Variable		Mean	Std. Dev.	Min	Max	Observations	
Housep~e	overall	101.2638	15.5443	63.88	168.0575	N =	600
	between		8.740909	88.9715	121.906	n =	15
	within		13.04593	74.32258	147.4152	T =	40
CPI	overall	1.239576	1.318307	-6.12777	5.533818	N =	600
	between		.4753793	.0577659	2.1175	n =	15
	within		1.235581	-4.94596	5.955927	T =	40
GDP	overall	213997.1	221528.9	9030.6	851231	N =	600
	between		227263.1	11847.77	722846.1	n =	15
	within		28094.16	96858.5	342382	T =	40
Popula~n	overall	26813.17	27096.89	492.98	82926	N =	600
	between		28019.71	548.957	81210.38	n =	15
	within		504.0323	24516.39	29045.39	T =	40
Unempl~e	overall	9.571833	5.395168	3.4	27.7	N =	600
	between		5.047922	5.2625	20.86	n =	15
	within		2.298996	-2.985667	16.71433	T =	40
Wages	overall	81519.02	88303.04	4042.3	361020	N =	600
	between		90563.2	5028.105	297330.9	n =	15
	within		11390.08	34623.12	145208.1	T =	40
Intere~e	overall	2.947033	2.96501	-.12	25.4	N =	600
	between		2.151186	1.49075	9.78325	n =	15
	within		2.113049	-2.766217	18.56378	T =	40

Source: own elaboration.

Our dependent variable is *house price index*, and the rest are independent variables. We consider 5% significance level in our models.

4.1 Stationarity of data

4.1.1 Unit root tests

From Chapter 3, we know that unit root tests are used to investigate stationarity of data that affects the regression model. We have already mentioned the basic idea of unit root tests. Here, we use Levin-Lin-Chu test to detect stationarity of our data.

Levin-Lin-Chu test is based on ADF test. The null hypothesis is $\alpha = 1$ or $\delta = 0$ according to Formula (3.1) and (3.2), which means panels contain unit roots. The alternative hypothesis is $\alpha < 1$ or $\delta < 0$, which means panels are stationary.

Based on the analysis in STATA, we get following results. Details are shown in Annexes.

Table 4.2 Summary of LLC test

Variable	P-value	Result
House price index	0.9941	Non-stationary
CPI	0.0000	Stationary
GDP	0.9953	Non-stationary
Population	0.0694	Non-stationary
Unemployment rate	0.7903	Non-stationary
Wages	1.0000	Non-stationary
Interest rate	0.0041	Stationary

Source: own elaboration.

In this test, we want to reject the null hypothesis and accept the alternative hypothesis. Hence, we need the p-value to be lower than 5%. According to this table, except *CPI* and *Interest rate*, the other p-values are higher than 5%. These variables are non-stationary.

We need stationary data to proceed our analysis.

4.1.2 Transformation

Because five variables of our model are non-stationary. We try to transform them to be stationary. Here, we use two ways.

We start with first-differences. The first-difference estimator is an approach used to focus on the problem of omitted variables in econometrics with panel data. The estimator is obtained by running a pooled OLS estimation for a regression of differenced variables.

Table 4.3 Summary of LLC test in first difference

Variables	P-value	Results
D.House price index	0.0000	Stationary
D.CPI	0.0000	Stationary
D.GDP	0.0000	Stationary
D.Population	0.0029	Stationary
D.Unemployment rate	0.0000	Stationary
D.Wages	0.0000	Stationary
D.Interest rate	0.0000	Stationary

Source: own elaboration.

We can see that all differenced variables are stationary now.

Next, we use growth rates of the data. We can calculate percentage change of variables compared to same period in previous year in STATA. In addition, we do not focus on *CPI* and *interest rate* because they are stationary. Therefore, we transform only five variables.

Table 4.4 Summary of LLC test in growth rate

Variables	Results
Growth rate of house price index	Stationary
Growth rate of GDP	Stationary
Growth rate of population	Stationary
Growth rate of unemployment rate	Stationary
Growth rate of wages	Stationary

Source: own elaboration.

Both of ways can be successful. We choose to use year-on-year growth rates of variables to continue our analysis, since it is easier for us to understand changes among them regardless of the impact of seasonality.

4.2 Model estimation

4.2.1 Correlation between variables

First, we can look at the correlations across variables in the original model.

Table 4.5 Correlation matrix of original model

```
. pwcorr Houseprice CPI GDP Population Unemploymentrate Wages Interestrate
```

	Housep~e	CPI	GDP	Popula~n	Unempl~e	Wages	Intere~e
Houseprice	1.0000						
CPI	0.0225	1.0000					
GDP	0.0154	0.0618	1.0000				
Population	0.0536	0.0606	0.9749	1.0000			
Unemployme~e	0.1873	-0.2490	-0.1432	0.0087	1.0000		
Wages	-0.0159	0.0715	0.9907	0.9494	-0.1755	1.0000	
Interestrate	0.1861	0.1135	-0.2093	-0.1181	0.5944	-0.2287	1.0000

Source: own elaboration.

From above table, we can see that correlation between *GDP* and *population* and correlation between *GDP* and *wages* are higher than 0.8. It means these independent variables are closely correlated with each others, which may lead to inaccurate results of regression.

Then, we can test the correlation after transformation.

Table 4.6 Correlation matrix after transformation

```
. pwcorr GR_Houseprice CPI GR_GDP GR_Population GR_Unemploymentrate GR_Wages Interestrate
```

	GR_Hou~e	CPI	GR_GDP	GR_Pop~n	GR_Une~e	GR_Wages	Intere~e
GR_Housepr~e	1.0000						
CPI	-0.0728	1.0000					
GR_GDP	0.5043	-0.0538	1.0000				
GR_Populat~n	0.4000	0.1418	0.3148	1.0000			
GR_Unemplo~e	-0.5446	0.1496	-0.4534	-0.0466	1.0000		
GR_Wages	0.5704	-0.0163	0.7716	0.3587	-0.5408	1.0000	
Interestrate	-0.5791	0.1135	-0.4041	-0.4259	0.5045	-0.5892	1.0000

Source: own elaboration.

We are able to say that correlation of our variables after the transformation are acceptable, and we can proceed to the estimation. From Table 4.6, we can find that *interest rate* is the most correlated with dependent variable among six independent variables. On the other hand, *CPI* is the least correlated with *house price growth*.

4.2.2 Lag tests

We start with pooled OLS to regress the model.

Table 4.7 Regression model

Number of obs	540	Prob > F	0.00
F (6, 533)	90.56	Adj R-squared	0.50
GR_House price index	Coefficient	t	P>| t |
CPI	-0.07	-0.48	0.631
GR_GDP	0.17	2.77	0.006
GR_Population	2.04	6.00	0.000
GR_Unemployment rate	-0.14	-7.86	0.000
GR_Wages	0.12	1.58	0.114
Interest rate	-0.42	-5.02	0.000
Constant	1.32	3.17	0.002

Source: own elaboration.

From this table, *number of observations* is 540 which is lower than 600, because we use variables by year-on-year growth rates. The periods of each country would be reduced by 4 and we have 15 countries. *F(6.533)* represents the result of F-test. The result of *Probability > F* shows the whole model is statistically significant. *Adjusted R-squared* measures the 0.5 variability in the *growth rate of house price index* that can be explained by the six explanatory variables. The other part of this table shows results of t-test. We need to detect if each coefficient is statistically significant.

We can find that p-values of CPI and growth rate of wages are higher than 5%, which means these two variables are insignificant, i.e. their coefficients are not significant different from zero. Therefore, these two variables do not influence *growth rate of house price index*.

However, we know at least *growth rate of wages* will certainly affect house prices. From the perspective of econometrics, we can find out the high correlation between *growth rate of wages* and *growth rate of house price index* according to Table 4.6. Also, we have already discussed relationship between wages and house prices in daily life in Chapter 2.

We can also detect if there is autocorrelation in this model, based on Formula (3.18).

Table 4.8 Autocorrelation

```
. pwcorr ut ut_1, sig
```

	ut	ut_1
ut	1.0000	
ut_1	0.9003 0.0000	1.0000

Source: own elaboration.

In this table, *ut* means residuals and *ut_1* refers to residuals which lose the first time period. We can find that these two variables are statistically significant, because p-value is equal to 0.0000. There is autocorrelation in the original model. We need to remove this autocorrelation.

We have already mentioned some ways in Chapter 3. In this part, we choose to consider the impact of time lag (L) on independent variables.

We know that not all changes in independent variables will immediately affect the dependent variable. For instance, when our wage increases, we will not buy house immediately. We are more likely to wait until the wages rise steadily for a while before we consider buying a house.

Actually, when we input different lags into variables, results are quite good-- there are a lot of models we can choose. We tested all alternatives and we choose three sets of lags that are shown below. The estimations based on pooled OLS are named Model 1, Model 2, and Model 3. Full results are available in Annexes.

In Model 1, there are no lags in *CPI*, *growth rate of GDP*, *growth rate of population* and *growth rate of unemployment rate*. There are four lags in *growth rate of wages* and *level of interest rate*. In Model 2, one lag is in *CPI* and *growth rate of unemployment rate*. No lags are in *growth rate of GDP* and *growth rate of population.* Six lags are in *growth rate of wages* and four lags are in *interest rate*. In Model 3, one lag is in *CPI* and *growth rate of GDP*. Two lags are in *growth rate of population* and *growth rate of unemployment rate*. Six lags are in *growth rate of wages* and four lags are in *interest rate*.

Table 4.9 Model 1 (pooled OLS model)

Number of obs	480	Prob > F	0.00
F (6, 473)	107.67	Adj R-squared	0.57
GR_House price index	Coefficient	t	P>| t |
CPI	-0.46	-2.89	0.004
GR_GDP	0.17	3.76	0.000
GR_Population	2.06	6.15	0.000
GR_Unemployment rate	-0.19	-9.41	0.000
L4. GR_Wages	0.21	4.12	0.000
L4. Interest rate	-0.32	-4.11	0.000
Constant	1.20	2.53	0.012

Table 4.10 Model 2 (pooled OLS model)

Number of obs	450	Prob > F	0.00
F (6, 443)	112.72	Adj R-squared	0.60
GR_House price index	Coefficient	t	P>| t |
L1. CPI	-0.66	-4.19	0.000
GR_GDP	0.19	4.22	0.000
GR_Population	2.10	6.29	0.000
L1. GR_Unemployment rate	-0.19	-9.70	0.000
L6. GR_Wages	0.27	5.35	0.000
L4. Interest rate	-0.22	-2.81	0.005
Constant	1.07	2.33	0.020

Table 4.11 Model 3 (pooled OLS model)

Number of obs	450	Prob > F	0.00
F (6, 443)	102.96	Adj R-squared	0.58
GR_House price index	Coefficient	t	P>| t |
L1. CPI	-0.83	-5.22	0.000
L1. GR_GDP	0.15	3.37	0.001
L2. GR_Population	1.87	5.40	0.000
L2. GR_Unemployment rate	-0.20	-9.38	0.000
L6. GR_Wages	0.25	4.79	0.000
L4. Interest rate	-0.24	-3.06	0.002
Constant	1.68	3.65	0.000

Source: own elaboration.

From above models, all three models have higher adjusted R-squared values than the original regression model. Because of different lags, models have different observations and R-squared. Model 2 has the highest R-squared value 0.60.

The effect of *CPI* on *house price index growth* ranges from -0.83 to -0.46. The effect of *GDP growth* ranges from 0.15 to 0.19. The effect of *population growth* ranges from 1.87 to 2.10. The effect of *unemployment rate growth* ranges from -0.20 to -0.19. The effect of *wages growth* ranges from 0.21 to 0.27. The effect of *interest rate* ranges from -0.32 to -0.22. They are all statistically significant at 5% level.

Then, we estimate Fixed Effect models with the same three sets of lags and we call them Model 4, Model 5, and Model 6. Full results are shown in Annexes.

Table 4.12 Model 4 (FE model)

Number of obs	480	Prob > F	0.00
F (6, 459)	93.16	R-squared: overall	0.51
GR_House price index	Coefficient	t	P>| t |
CPI	-0.63	-4.10	0.000
GR_GDP	0.20	4.95	0.000
GR_Population	5.98	8.02	0.000
GR_Unemployment rate	-0.19	-9.66	0.000
L4. GR_Wages	0.14	3.08	0.002
L4. Interest rate	-0.35	-3.61	0.000
Constant	-0.46	-0.75	0.452

Table 4.13 Model 5 (FE model)

Number of obs	450	Prob > F	0.00
F (6, 429)	92.74	R-squared: overall	0.55
GR_House price index	Coefficient	t	P>| t |
L1. CPI	-0.84	-5.26	0.000
GR_GDP	0.21	4.98	0.000
GR_Population	5.41	7.05	0.000
L1. GR_Unemployment rate	-0.18	-9.06	0.000
L6. GR_Wages	0.20	4.38	0.000
L4. Interest rate	-0.23	-2.22	0.027
Constant	-0.27	-0.44	0.662

Table 4.14 Model 6 (FE model)

Number of obs	450	Prob > F	0.00
F (6, 443)	77.18	R-squared: overall	0.57
GR_House price index	Coefficient	t	P>| t |
L1. CPI	-1.04	-6.24	0.000
L1. GR_GDP	0.16	3.55	0.000
L2. GR_Population	3.29	4.14	0.000
L2. GR_Unemployment rate	-0.18	-8.60	0.000
L6. GR_Wages	0.19	3.91	0.000
L4. Interest rate	-0.29	-2.61	0.009
Constant	1.47	2.34	0.020

Source: own elaboration.

Model 6 has the highest R-squared value among these three models.

In these models, the effect of *CPI* on *house price index growth* ranges from -1.04 to -0.63. The effect of *GDP growth* ranges from 0.16 to 0.21. The effect of *population growth* ranges from 3.29 to 5.98. The effect of *unemployment rate growth* ranges from -0.19 to -0.18. The effect of *wages growth* ranges from 0.14 to 0.20. The effect of *interest rate* ranges from -0.35 to -0.23. They are all statistically significant at 5% level.

4.2.3 The selected model

All six models provide reasonable results. The best performing model seems to be Model 6, because all independent variables have lags in this model, and Fixed Effect model is better to conduct the individual-specific effects.

In Model 6, it has the same set of lags as Model 3. There is one lag is in *CPI* and *GDP growth*. There are two lags are in *population growth* and *unemployment rate growth*. Six lags are in *wages growth* and four lags are in *interest rate*. These lags are quite reasonable in the reality. In reality, it usually takes some time before the change in independent variables influence *house prices growth*.

Householders are not so forward looking. They do not make too much predictions of *CPI* and *growth of GDP*, but they focus more on what happens now or what was happening a quarter ago. *GDP growth* and GDP prediction, especially, are usually reported in the news. There is no need for people to consider the situation of their country's GDP. News will tell them if GDP is expected to increase or not. They can make choices if they want to buy houses now or a litter bit later. The central banks nowadays informs the households about price level developments and they can reflect this information into their demand for properties soon. Hence, lags of these two variables are small.

Two lags are in *population growth.* There are many reasons for *population growth* in a country. For example, when a newborn baby is born, parents are more likely to move from a small apartment to a bigger property. But when the child is born, the family could not immediately buy a house and move into there. This process takes a certain amount of time. Another example is the increase in immigrants. When they arrive in a new country, they need to know a lot of information and handle a lot of things before buying houses. However, when they have already settled down, they need a property immediately. Hence, the effect of *population growth* on the *house prices growth* is delayed but time lags are not so long.

Also, two lags are in *unemployment rate growth. C*ompared to the effect of *GDP growth and CPI, unemployment rate growth* usually takes a bit more time to transfer the its effect to the *growth of house prices*.

If the *wages growth* is small, it takes sometime to accumulate wealth. Like we discussed before, when our wages just start to increase, we will not buy a house in the market immediately. We are more likely to wait for a steadily increase lasting for one or two years.

The same as interest rate, when interest rate falls, most people will wait for a while if it falls further or not.

We can also detect if there is multicollinearity in this model. We provide correlation matrix of Model 6 in the table below.

Table 4.15 Correlation matrix of Model 6

```
. pwcorr GR_Houseprice L.CPI L.GR_GDP L2.GR_Population L2.GR_Unemploymentrate L6.GR_Wages L4.Interestrate
```

	GR_Hou~e	L.CPI	L.GR_GDP	L2.GR_~n	L2.G~U~e	L6.GR_~s	L4.Int~e
GR_Housepr~e	1.0000						
L.CPI	-0.0873	1.0000					
L.GR_GDP	0.4724	-0.0541	1.0000				
L2.GR_Popu~n	0.3652	0.1679	0.3097	1.0000			
L2.GR_Unem~e	-0.5283	0.1866	-0.3847	-0.0564	1.0000		
L6.GR_Wages	0.4740	0.1211	0.1989	0.4421	-0.2983	1.0000	
L4.Interes~e	-0.5572	-0.0051	-0.3715	-0.4566	0.4570	-0.5566	1.0000

Source: own elaboration.

All values of correlation between independent variables are lower than 0.8. There is no multicollinearity in this model. Compared to Table 4.6, wages growth in six lags are lower correlated with GDP growth in one lag.

We can easily order correlation values of independent variables on dependent variables. *Interest rate* has the highest correlation value with *house price growth*.

When variables are measured in different units of measurement, the standardized coefficients are usually performed to answer the question of which independent variable has a greater effect on the dependent variable in the multiple regression analysis. For example, in our models, units of some variables are growth rate but units of others are level of index. Each variable can be standardized by subtracting its mean from each of its values and then dividing these new values by the standard deviation of the variables.

After computation of STATA, we can get standardized Model 6 below. Full result is shown in Annexes.

Table 4.16 Standardized Model 6

zGR_House price index	Coefficient	t	P>| t |
L1. zCPI	-0.23	-6.24	0.000
L1. zGR_GDP	0.13	3.55	0.000
L2. zGR_Population	0.36	4.14	0.000
L2. zGR_Unemployment rate	-0.41	-8.60	0.000
L6.zGR_Wages	0.14	3.91	0.000
L4. zInterest rate	-0.15	-2.61	0.009
Constant	0.01	0.32	0.752

Source: own elaboration.

The standardized table focuses on absolute value of coefficients.

Growth rate of unemployment rate has the highest absolute value and *growth rate of population* has the second highest absolute value. The strongest reaction on *house price index growth* is from these two variables basically. When people lose jobs, they definitely do not plan to buy a house. They can only try to survive on the basic level of living condition. People need to live in a property. When *population growth* increases, *house prices growth* is more likely to react.

GDP growth has the lowest absolute value. Because even in some rich countries, there are people who cannot afford a property.

4.3 Summary of models

We have already provided results for six selected models. Then, we can provide the summary of the estimation results in Table 4.17.

Table 4.17 Summary of estimation results

	CPI	GR_GDP	GR_Population	GR_Unemployment rate	GR_Wages	Interest rate
Pooled OLS models						
M1	−	+	+	−	+	−
M2	−	+	+	−	+	−
M3	−	+	+	−	+	−
FE models						
M4	−	+	+	−	+	−
M5	−	+	+	−	+	−
M6	−	+	+	−	+	−

Source: own elaboration.

We can find that all six models are acceptable, and signs of all independent variables are the same in six models.

CPI is negatively correlated with our dependent variable. This might be still in line with economic theory. When *CPI* is higher, people have to spend more on ordinary goods and services. Therefore, they do not have enough money to afford a house. It reduces their purchasing power and reduces the demand of property. *Growth of house prices* will decrease. This is the only variable which is different with what we suggested in the end of Chapter 2. We assume *CPI* is positively correlated with our dependent variable.

There is a positive correlation between *growth rate of GDP* and *growth rate of house price*. It is reasonable. Usually, when a country is richer, people in this country have more money to buy a property.

The effect of *growth rate of population* on *growth of house prices* is positive which is in line with our suggestion. When it increases, people will purchase more houses and *house prices growth* will increase.

Growth rate of unemployment rate is negatively correlated with our dependent variable. When people lose jobs, they do not have incomes and they cannot spend much money on houses. *House prices growth* will therefore decrease.

Growth rate of wages are the opposite of *growth rate of unemployment rate*.

Interest rate is negatively correlated with *growth rate of house price index*. When *interest rate* is higher, it is harder for people to repay interest of loan and fewer people can afford to buy a house.

5. Conclusion

In this thesis, we analyze the effect of macroeconomic determinants on property prices. We use panel data regression model to study the influence of selected six variables on property prices for 15 European countries from the fourth quarter of 2008 to the third quarter of 2018.

In the introduction, we explain why our research is relevant and up to date. Our main argument is based on motivation of households and corporations for property purchases and financial stability motive of central banks. In the second chapter, we briefly introduce the participants and characteristics of the property market, and provide theoretically review of the determinants of housing prices. We focus on macroeconomic variables, such as CPI, GDP, population, unemployment rate and so on. In the third chapter, we theoretically explain the panel data model and introduce the pooled OLS and fixed effect methods. In Chapter 4, we empirically examine the stationarity of data and estimate acceptable models, then we make a summary of our six preferred models. These models where selected based on statistically significance, explanatory power and consistency with economic theory.

Model 6 is our preferred model which there is one lag in CPI and GDP growth. Two lags are in population growth and unemployment rate growth. Six lags are in wages growth and four lags are in interest rate. These lags are quite reasonable in the reality. People usually do not decide based on forecasting actual values of these variables when they need to buy a property. Even if they do so, it takes some time before the intention changes into reality.

Also, we make a summary of estimation results. We can find that six independent variables have different effects on dependent variable and the results are robust. CPI has negative effect on house price growth. The effect of GDP growth and population growth on house prices growth are positive. Unemployment rate growth is negatively correlated with house prices growth. Wages growth has the opposite effect comparing with unemployment rate growth. Interest rate shows negative effect on dependent variable.

We can conclude that housing prices are pro-cyclical and there may be conflict between monetary policy and financial stability. Further, conflicts may arise from proactive employment policy. Pressures on salary growth may accelerate property price growth.

We hope this thesis can provide a reference for researchers who study volatility of property prices and determinants on property market. We hope that investors and household can get favorable information from this thesis. Our model may also provide useful information for financial stability departments in central banks.

For further research, we would like to apply GMM (generalized method of moments) or country specific models. Further, we plan to enlarge the model by adding more variables from the supply side.

Bibliography

Professional books

[1] BROOKS, Chris and Sotiris TSOLACOS. *Real Estate Modelling and Forecasting.* 1st ed. Cambridge: Cambridge University Press, 2010. 474 p. ISBN: 9780521873390.

[2] DE BANDT, Oliver and Thomas KNETSCH. *Housing Markets in Europe: A Macroeconomic Perspective.* 1st ed. Berlin: Springer, 2010. 406 p. ISBN: 9783642153396.

[3] GUJARATI, Damodar N. *Basic Econometrics.* 4th ed. New York: McGraw-Hill, 2003. 1002 p. ISBN: 9780072335422.

[4] MADDALA, G.S. *Introduction to Econometrics*. 3rd ed. New York: Wiley, 2001. 664 p. ISBN: 9780471497288.

[5] WOOLDRIDGE, Jeffrey M. *Econometric Analysis of Cross Section and Panel Data.* 1st ed. MIT Press, 2002. 752 p. ISBN: 9780262232197

Electronic Bibliography

[6] BILOZOR, M. Renigier. *The Impact of Macroeconomic Factors on Residential Property Price Indices in Europe.* [online] [30.07.2013] Available on https://content.sciendo.com/view/journals/foli/12/2/article-p103.xml

[7] CASE, Karl E. and Robert J. SHILLER. *The Efficiency of the Market for Single-Family Homes.* [online] [02.1988]. Available on https://www.nber.org/papers/w2506

[8] ENGLUND, Peter and Yannis M. IOANNIDES. *House Price Dynamics: An International Empirical Perspective.* [online] [06.1997] Available on https://www.sciencedirect.com/science/article/pii/S1051137797902102#!

[9] European Mortgage Federation [online] [17.01.2019] Available on: https://hypo.org/emf/press-release/emf-publishes-quarterly-review-q3-2018/

[10] Eurostat [online] [02. 2019]. Available on https://ec.europa.eu/eurostat/data/database

[11] FOLLAIN, James R. *Does inflation affect real behavior? The case of housing*. [online] [01.1982]. Available on https://www.jstor.org/stable/1058651?seq=1#page_scan_tab_contents

[12] GENESOVE, David and Christopher MAYER. *The Quarterly Journal of Economics.*

[online] [11. 2001]. Available on https://doi.org/10.1162/003355301753265561

[13] HEJLOVA, H., M. HLAVACEK and L.KOMAREK. *A Comprehensive Method for House Price Sustainability Assessment in the Czech Republic.* [online] [2017]. Available on https://ideas.repec.org/a/prg/jnlpep/v2017y2017i3id613p269-285.html

[14] LAMONT, Owen A. *Economic Tracking Portfolios.* [online] [23.03.1999]. Available on https://ssrn.com/abstract=157772

[15] OECD [online] [02.2019]. Available on https://data.oecd.org/

[16] Oxford English Dictionary [online] [18.09.2011]. Available on https://en.oxforddictionaries.com/definition/real_estate

List of Abbreviations

CPI	Consumer price index
GDP	Gross domestic product
ADF	Augmented Dickey-Fuller
DF	Dickey-Fuller
KPSS	Kwiatkowski–Phillips–Schmidt–Shin
OLS	Ordinary least squares
FE	Fixed effects
RE	Random effects
SSR	Sum of squared residuals
VIF	Variance inflation factor
JB	Jarque–Bera
LLC	Levin-Lin-Chu
GMM	Generalized method of moments

Declaration of Utilisation of Results from the Diploma Thesis

Herewith I declare that

— I am informed that Act No. 121/2000 Coll. – the Copyright Act, in particular, Section 35 – Utilisation of the Work as a Part of Civil and Religious Ceremonies, as a Part of School Performances and the Utilisation of a School Work – and Section 60 – School Work, fully applies to my diploma thesis;

— I take account of the VSB – Technical University of Ostrava (hereinafter as VSB-TUO) having the right to utilize the diploma thesis (under Section 35(3)) unprofitably and for own use ;

— I agree that the diploma thesis shall be archived in the electronic form in VSB-TUO's Central Library. I agree that the bibliographic information about the diploma thesis shall be published in VSB-TUO's information system;

— It was agreed that, in case of VSB-TUO's interest, I shall enter into a license agreement with VSB-TUO, granting the authorization to utilize the work in the scope of Section 12(4) of the Copyright Act;

— It was agreed that I may utilize my work, the diploma thesis or provide a license to utilize it only with the consent of VSB-TUO, which is entitled, in such a case, to claim an adequate contribution from me to cover the cost expended by VSB-TUO for producing the work (up to its real amount).

Ostrava dated 22. 4. 2019

Siyu Liu 刘思雨

Student's name and surname

List of Annexes

Annex 1 Levin-Lin-Chu test

```
. xtunitroot llc Houseprice, lags(1)

Levin-Lin-Chu unit-root test for Houseprice
────────────────────────────────────────────
Ho: Panels contain unit roots              Number of panels  =     15
Ha: Panels are stationary                  Number of periods =     40

AR parameter: Common                       Asymptotics: N/T -> 0
Panel means:  Included
Time trend:   Not included

ADF regressions: 1 lag
LR variance:     Bartlett kernel, 10.00 lags average (chosen by LLC)
──────────────────────────────────────────────────────────────────────────────
                    Statistic      p-value
──────────────────────────────────────────────────────────────────────────────
 Unadjusted t        -0.9867
 Adjusted t*          2.5177        0.9941
──────────────────────────────────────────────────────────────────────────────
```

```
. xtunitroot llc CPI, lags(1)

Levin-Lin-Chu unit-root test for CPI
────────────────────────────────────
Ho: Panels contain unit roots              Number of panels  =     15
Ha: Panels are stationary                  Number of periods =     40

AR parameter: Common                       Asymptotics: N/T -> 0
Panel means:  Included
Time trend:   Not included

ADF regressions: 1 lag
LR variance:     Bartlett kernel, 10.00 lags average (chosen by LLC)
──────────────────────────────────────────────────────────────────────────────
                    Statistic      p-value
──────────────────────────────────────────────────────────────────────────────
 Unadjusted t       -10.1417
 Adjusted t*         -3.9967        0.0000
──────────────────────────────────────────────────────────────────────────────
```

```
. xtunitroot llc GDP, lags(1)

Levin-Lin-Chu unit-root test for GDP
------------------------------------
Ho: Panels contain unit roots              Number of panels  =     15
Ha: Panels are stationary                  Number of periods =     40

AR parameter: Common                       Asymptotics: N/T -> 0
Panel means:  Included
Time trend:   Not included

ADF regressions: 1 lag
LR variance:     Bartlett kernel, 10.00 lags average (chosen by LLC)
------------------------------------------------------------------------------
                   Statistic      p-value
------------------------------------------------------------------------------
 Unadjusted t        0.2369
 Adjusted t*         2.5951        0.9953
------------------------------------------------------------------------------


. xtunitroot llc Population, lags(1)

Levin-Lin-Chu unit-root test for Population
-------------------------------------------
Ho: Panels contain unit roots              Number of panels  =     15
Ha: Panels are stationary                  Number of periods =     40

AR parameter: Common                       Asymptotics: N/T -> 0
Panel means:  Included
Time trend:   Not included

ADF regressions: 1 lag
LR variance:     Bartlett kernel, 10.00 lags average (chosen by LLC)
------------------------------------------------------------------------------
                   Statistic      p-value
------------------------------------------------------------------------------
 Unadjusted t       -2.2050
 Adjusted t*        -1.4804        0.0694
------------------------------------------------------------------------------


. xtunitroot llc Unemploymentrate, lags(1)

Levin-Lin-Chu unit-root test for Unemploymentrate
-------------------------------------------------
Ho: Panels contain unit roots              Number of panels  =     15
Ha: Panels are stationary                  Number of periods =     40

AR parameter: Common                       Asymptotics: N/T -> 0
Panel means:  Included
Time trend:   Not included

ADF regressions: 1 lag
LR variance:     Bartlett kernel, 10.00 lags average (chosen by LLC)
------------------------------------------------------------------------------
                   Statistic      p-value
------------------------------------------------------------------------------
 Unadjusted t       -4.1013
 Adjusted t*         0.8074        0.7903
```

```
. xtunitroot llc Wages, lags(1)

Levin-Lin-Chu unit-root test for Wages
--------------------------------------
Ho: Panels contain unit roots           Number of panels  =     15
Ha: Panels are stationary               Number of periods =     40

AR parameter: Common                    Asymptotics: N/T -> 0
Panel means:  Included
Time trend:   Not included

ADF regressions: 1 lag
LR variance:     Bartlett kernel, 10.00 lags average (chosen by LLC)
------------------------------------------------------------------------------
                   Statistic      p-value
------------------------------------------------------------------------------
 Unadjusted t         3.9153
 Adjusted t*          6.7755       1.0000
------------------------------------------------------------------------------
```

```
. xtunitroot llc Interestrate, lags(1)

Levin-Lin-Chu unit-root test for Interestrate
---------------------------------------------
Ho: Panels contain unit roots           Number of panels  =     15
Ha: Panels are stationary               Number of periods =     40

AR parameter: Common                    Asymptotics: N/T -> 0
Panel means:  Included
Time trend:   Not included

ADF regressions: 1 lag
LR variance:     Bartlett kernel, 10.00 lags average (chosen by LLC)
------------------------------------------------------------------------------
                   Statistic      p-value
------------------------------------------------------------------------------
 Unadjusted t        -5.4033
 Adjusted t*         -2.6462       0.0041
------------------------------------------------------------------------------
```

Annex 2 Levin-Lin-Chu test in first difference

```
. xtunitroot llc D.Houseprice, lags(1)

Levin-Lin-Chu unit-root test for D.Houseprice
---------------------------------------------
Ho: Panels contain unit roots            Number of panels  =     15
Ha: Panels are stationary                Number of periods =     39

AR parameter: Common                     Asymptotics: N/T -> 0
Panel means:  Included
Time trend:   Not included

ADF regressions: 1 lag
LR variance:     Bartlett kernel, 10.00 lags average (chosen by LLC)
------------------------------------------------------------------------------
                    Statistic      p-value
------------------------------------------------------------------------------
 Unadjusted t        -9.6108
 Adjusted t*         -4.2871        0.0000
------------------------------------------------------------------------------

. xtunitroot llc D.GDP, lags(1)

Levin-Lin-Chu unit-root test for D.GDP
--------------------------------------
Ho: Panels contain unit roots            Number of panels  =     15
Ha: Panels are stationary                Number of periods =     39

AR parameter: Common                     Asymptotics: N/T -> 0
Panel means:  Included
Time trend:   Not included

ADF regressions: 1 lag
LR variance:     Bartlett kernel, 10.00 lags average (chosen by LLC)
------------------------------------------------------------------------------
                    Statistic      p-value
------------------------------------------------------------------------------
 Unadjusted t       -14.8158
 Adjusted t*         -5.3694        0.0000
------------------------------------------------------------------------------

. xtunitroot llc D.Population, lags(1)

Levin-Lin-Chu unit-root test for D.Population
---------------------------------------------
Ho: Panels contain unit roots            Number of panels  =     15
Ha: Panels are stationary                Number of periods =     39

AR parameter: Common                     Asymptotics: N/T -> 0
Panel means:  Included
Time trend:   Not included

ADF regressions: 1 lag
LR variance:     Bartlett kernel, 10.00 lags average (chosen by LLC)
------------------------------------------------------------------------------
                    Statistic      p-value
------------------------------------------------------------------------------
 Unadjusted t        -6.9564
 Adjusted t*         -2.7563        0.0029
```

```
. xtunitroot llc D.Unemploymentrate, lags(1)

Levin-Lin-Chu unit-root test for D.Unemploymentrate
---------------------------------------------------
Ho: Panels contain unit roots           Number of panels  =     15
Ha: Panels are stationary               Number of periods =     39

AR parameter: Common                    Asymptotics: N/T -> 0
Panel means:  Included
Time trend:   Not included

ADF regressions: 1 lag
LR variance:     Bartlett kernel, 10.00 lags average (chosen by LLC)
------------------------------------------------------------------------------
                  Statistic      p-value
------------------------------------------------------------------------------
 Unadjusted t      -12.2112
 Adjusted t*        -6.2991       0.0000
```

```
. xtunitroot llc D.Wages, lags(1)

Levin-Lin-Chu unit-root test for D.Wages
----------------------------------------
Ho: Panels contain unit roots           Number of panels  =     15
Ha: Panels are stationary               Number of periods =     39

AR parameter: Common                    Asymptotics: N/T -> 0
Panel means:  Included
Time trend:   Not included

ADF regressions: 1 lag
LR variance:     Bartlett kernel, 10.00 lags average (chosen by LLC)
------------------------------------------------------------------------------
                  Statistic      p-value
------------------------------------------------------------------------------
 Unadjusted t      -12.3357
 Adjusted t*        -5.2417       0.0000
------------------------------------------------------------------------------
```

Annex 3 Levin-Lin-Chu test in growth rate

```
. gen GR_Houseprice=S4.Houseprice/L4.Houseprice *100
(60 missing values generated)

. gen GR_GDP=S4.GDP/L4.GDP *100
(60 missing values generated)

. gen GR_Population=S4.Population/L4.Population *100
(60 missing values generated)

. gen GR_Unemploymentrate =S4.Unemploymentrate /L4.Unemploymentrate *100
(60 missing values generated)

. gen GR_Wages =S4.Wages /L4.Wages *100
(60 missing values generated)
```

```
Levin-Lin-Chu unit-root test for GR_Houseprice
----------------------------------------------
Ho: Panels contain unit roots             Number of panels  =     15
Ha: Panels are stationary                 Number of periods =     36

AR parameter: Common                      Asymptotics: N/T -> 0
Panel means:  Included
Time trend:   Not included

ADF regressions: 1 lag
LR variance:     Bartlett kernel, 10.00 lags average (chosen by LLC)
------------------------------------------------------------------------------
                    Statistic      p-value
------------------------------------------------------------------------------
 Unadjusted t        -6.5873
 Adjusted t*         -1.9045        0.0284
------------------------------------------------------------------------------
```

```
Levin-Lin-Chu unit-root test for GR_GDP
---------------------------------------
Ho: Panels contain unit roots             Number of panels  =     15
Ha: Panels are stationary                 Number of periods =     36

AR parameter: Common                      Asymptotics: N/T -> 0
Panel means:  Included
Time trend:   Not included

ADF regressions: 1 lag
LR variance:     Bartlett kernel, 10.00 lags average (chosen by LLC)
------------------------------------------------------------------------------
                    Statistic      p-value
------------------------------------------------------------------------------
 Unadjusted t        -8.0387
 Adjusted t*         -1.7590        0.0393
------------------------------------------------------------------------------
```

```
Levin-Lin-Chu unit-root test for GR_Population
----------------------------------------------
Ho: Panels contain unit roots           Number of panels  =     15
Ha: Panels are stationary               Number of periods =     36

AR parameter: Common                    Asymptotics: N/T -> 0
Panel means:  Included
Time trend:   Included

ADF regressions: 1 lag
LR variance:     Bartlett kernel, 10.00 lags average (chosen by LLC)
------------------------------------------------------------------------------
                   Statistic      p-value
------------------------------------------------------------------------------
 Unadjusted t        -8.6794
 Adjusted t*         -3.8768        0.0001
------------------------------------------------------------------------------

Levin-Lin-Chu unit-root test for GR_Unemploymentrate
----------------------------------------------------
Ho: Panels contain unit roots           Number of panels  =     15
Ha: Panels are stationary               Number of periods =     36

AR parameter: Common                    Asymptotics: N/T -> 0
Panel means:  Included
Time trend:   Included                  Cross-sectional means removed

ADF regressions: 1 lag
LR variance:     Bartlett kernel, 10.00 lags average (chosen by LLC)
------------------------------------------------------------------------------
                   Statistic      p-value
------------------------------------------------------------------------------
 Unadjusted t        -9.5323
 Adjusted t*         -1.8674        0.0309
------------------------------------------------------------------------------

Levin-Lin-Chu unit-root test for GR_Wages
-----------------------------------------
Ho: Panels contain unit roots           Number of panels  =     15
Ha: Panels are stationary               Number of periods =     36

AR parameter: Common                    Asymptotics: N/T -> 0
Panel means:  Included
Time trend:   Not included

ADF regressions: 1 lag
LR variance:     Bartlett kernel, 10.00 lags average (chosen by LLC)
------------------------------------------------------------------------------
                   Statistic      p-value
------------------------------------------------------------------------------
 Unadjusted t        -7.9980
 Adjusted t*         -1.7362        0.0413
------------------------------------------------------------------------------
```

Annex 4 Original regression model

```
. reg GR_Houseprice CPI GR_GDP GR_Population GR_Unemploymentrate GR_Wages Interestrate
```

Source	SS	df	MS
Model	9430.26118	6	1571.7102
Residual	9250.23206	533	17.355032
Total	18680.4932	539	34.6576869

Number of obs	=	540
F(6, 533)	=	90.56
Prob > F	=	0.0000
R-squared	=	0.5048
Adj R-squared	=	0.4992
Root MSE	=	4.1659

GR_Houseprice	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
CPI	-.0724555	.1508154	-0.48	0.631	-.368721	.2238101
GR_GDP	.1660877	.0599897	2.77	0.006	.0482424	.283933
GR_Population	2.035417	.339121	6.00	0.000	1.369239	2.701594
GR_Unemploymentrate	-.1376579	.017517	-7.86	0.000	-.1720687	-.1032471
GR_Wages	.1252459	.0790549	1.58	0.114	-.0300515	.2805432
Interestrate	-.4150008	.0827474	-5.02	0.000	-.5775519	-.2524497
_cons	1.313295	.4147221	3.17	0.002	.4986042	2.127985

Annex 5 Three pooled OLS models

```
. reg GR_Houseprice CPI GR_GDP GR_Population GR_Unemploymentrate L4.GR_Wages L4.Interestrate
```

Source	SS	df	MS			
				Number of obs	=	480
				F(6, 473)	=	107.67
Model	9462.12064	6	1577.02011	Prob > F	=	0.0000
Residual	6928.04326	473	14.6470259	R-squared	=	0.5773
				Adj R-squared	=	0.5719
Total	16390.1639	479	34.2174612	Root MSE	=	3.8271

GR_Houseprice	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
CPI	-.4565183	.1577981	-2.89	0.004	-.7665904	-.1464463
GR_GDP	.1660126	.044163	3.76	0.000	.0792326	.2527925
GR_Population	2.060763	.3351893	6.15	0.000	1.402119	2.719407
GR_Unemploymentrate	-.1886888	.0200488	-9.41	0.000	-.2280844	-.1492931
GR_Wages						
L4.	.2126701	.0515897	4.12	0.000	.1112967	.3140435
Interestrate						
L4.	-.3164603	.0770197	-4.11	0.000	-.4678034	-.1651172
_cons	1.197231	.472706	2.53	0.012	.2683673	2.126094

```
. reg GR_Houseprice L.CPI GR_GDP GR_Population L.GR_Unemploymentrate L6.GR_Wages L4.Interestrate
```

Source	SS	df	MS			
				Number of obs	=	450
				F(6, 443)	=	112.72
Model	9310.39388	6	1551.73231	Prob > F	=	0.0000
Residual	6098.25479	443	13.7658122	R-squared	=	0.6042
				Adj R-squared	=	0.5989
Total	15408.6487	449	34.3177031	Root MSE	=	3.7102

GR_Houseprice	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
CPI						
L1.	-.6563284	.1567427	-4.19	0.000	-.96438	-.3482768
GR_GDP	.1913722	.0453207	4.22	0.000	.1023019	.2804425
GR_Population	2.096057	.3334154	6.29	0.000	1.440784	2.751329
GR_Unemploymentrate						
L1.	-.1940274	.0199981	-9.70	0.000	-.2333303	-.1547246
GR_Wages						
L6.	.265298	.0496022	5.35	0.000	.1678131	.3627828
Interestrate						
L4.	-.2158152	.0768267	-2.81	0.005	-.3668053	-.0648252
_cons	1.066954	.4586456	2.33	0.020	.165563	1.968346

. reg GR_Houseprice L.CPI L.GR_GDP L2.GR_Population L2.GR_Unemploymentrate L6.GR_Wages L4.Interestrate

Source	SS	df	MS			
				Number of obs	=	450
				F(6, 443)	=	102.96
Model	8973.44542	6	1495.57424	Prob > F	=	0.0000
Residual	6435.20325	443	14.5264182	R-squared	=	0.5824
				Adj R-squared	=	0.5767
Total	15408.6487	449	34.3177031	Root MSE	=	3.8114

GR_Houseprice	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
CPI						
L1.	-.8270664	.1583731	-5.22	0.000	-1.138322	-.5158105
GR_GDP						
L1.	.1539789	.0457395	3.37	0.001	.0640856	.2438722
GR_Population						
L2.	1.874482	.3472143	5.40	0.000	1.19209	2.556874
GR_Unemploymentrate						
L2.	-.1952788	.0208126	-9.38	0.000	-.2361825	-.1543751
GR_Wages						
L6.	.2457738	.051338	4.79	0.000	.1448776	.3466701
Interestrate						
L4.	-.2425729	.0793903	-3.06	0.002	-.3986014	-.0865445
_cons	1.681392	.4600377	3.65	0.000	.7772643	2.585519

Annex 6 Three FE models

```
. xtreg GR_Houseprice CPI GR_GDP GR_Population GR_Unemploymentrate L4.GR_Wages L4.Interestrate,fe

Fixed-effects (within) regression               Number of obs      =       480
Group variable: countrynum                      Number of groups   =        15

R-sq:                                           Obs per group:
     within  = 0.5491                                         min =        32
     between = 0.6635                                         avg =      32.0
     overall = 0.5133                                         max =        32

                                                F(6,459)           =     93.16
corr(u_i, Xb)  = -0.6635                        Prob > F           =    0.0000
```

GR_Houseprice	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
CPI	-.6320791	.1540115	-4.10	0.000	-.9347342	-.329424
GR_GDP	.2052055	.0414841	4.95	0.000	.1236832	.2867278
GR_Population	5.983303	.7456561	8.02	0.000	4.51798	7.448626
GR_Unemploymentrate	-.185829	.0192333	-9.66	0.000	-.2236253	-.1480328
GR_Wages						
L4.	.1466035	.047548	3.08	0.002	.0531649	.2400422
Interestrate						
L4.	-.3455239	.0957779	-3.61	0.000	-.5337415	-.1573063
_cons	-.4613122	.6126356	-0.75	0.452	-1.66523	.7426062
sigma_u	3.2194583					
sigma_e	3.4228295					
rho	.46941104	(fraction of variance due to u_i)				

F test that all u_i=0: F(14, 459) = 9.45 Prob > F = 0.0000

```
. xtreg GR_Houseprice L.CPI GR_GDP GR_Population L.GR_Unemploymentrate L6.GR_Wages L4.Interestrate,fe

Fixed-effects (within) regression               Number of obs      =       450
Group variable: countrynum                      Number of groups   =        15

R-sq:                                           Obs per group:
     within  = 0.5647                                          min =        30
     between = 0.6820                                          avg =      30.0
     overall = 0.5487                                          max =        30

                                                F(6,429)           =     92.74
corr(u_i, Xb)  = -0.5962                        Prob > F           =    0.0000

------------------------------------------------------------------------------------
      GR_Houseprice |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
--------------------+---------------------------------------------------------------
                CPI |
                L1. |  -.8395497   .1595334    -5.26   0.000    -1.153114   -.5259854
                    |
             GR_GDP |   .2178591   .0437493     4.98   0.000     .1318695    .3038488
      GR_Population |   5.409737   .7668638     7.05   0.000      3.90246    6.917015
                    |
GR_Unemploymentrate |
                L1. |  -.1812019   .0200009    -9.06   0.000    -.2205139   -.1418899
                    |
           GR_Wages |
                L6. |   .2032529   .0463604     4.38   0.000      .112131    .2943747
                    |
       Interestrate |
                L4. |  -.2322863    .104854    -2.22   0.027    -.4383779   -.0261948
                    |
              _cons |  -.2707029    .619556    -0.44   0.662    -1.488446      .94704
--------------------+---------------------------------------------------------------
            sigma_u |  2.8167758
            sigma_e |  3.3469928
                rho |  .41461031   (fraction of variance due to u_i)
------------------------------------------------------------------------------------
F test that all u_i=0: F(14, 429) = 8.24                     Prob > F = 0.0000
```

```
. xtreg GR_Houseprice L.CPI L.GR_GDP L2.GR_Population L2.GR_Unemploymentrate L6.GR_Wages L4.Interestrate, fe

Fixed-effects (within) regression               Number of obs      =       450
Group variable: countrynum                      Number of groups   =        15

R-sq:                                           Obs per group:
     within  = 0.5191                                         min =        30
     between = 0.7177                                         avg =      30.0
     overall = 0.5661                                         max =        30

                                                F(6,429)           =     77.18
corr(u_i, Xb)  = -0.2950                        Prob > F           =    0.0000

------------------------------------------------------------------------------------
       GR_Houseprice |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
---------------------+--------------------------------------------------------------
                 CPI |
                 L1. |  -1.037351   .1662758    -6.24   0.000    -1.364168   -.7105347
                     |
              GR_GDP |
                 L1. |   .1594861    .044936     3.55   0.000      .071164    .2478083
                     |
       GR_Population |
                 L2. |   3.286816   .7944229     4.14   0.000     1.725371    4.848262
                     |
 GR_Unemploymentrate |
                 L2. |  -.1824018   .0212007    -8.60   0.000     -.224072   -.1407315
                     |
            GR_Wages |
                 L6. |   .1915771   .0489654     3.91   0.000     .0953351     .287819
                     |
        Interestrate |
                 L4. |  -.2913037   .1117178    -2.61   0.009    -.5108862   -.0717213
                     |
               _cons |   1.473724   .6287141     2.34   0.020     .2379802    2.709467
---------------------+--------------------------------------------------------------
             sigma_u |  1.8946209
             sigma_e |   3.517836
                 rho |  .22484451   (fraction of variance due to u_i)
------------------------------------------------------------------------------------
F test that all u_i=0: F(14, 429) = 6.50                     Prob > F = 0.0000
```

Annex 7 Standardized Model 6

```
. xtreg zGR_Housepriceindex l.zCPI l.zGR_GDP l2.zGD_Population l2.zGD_Unemploymentrate l6.zGD_Wages l4.zInterestrate,fe

Fixed-effects (within) regression               Number of obs      =       450
Group variable: countrynum                      Number of groups   =        15

R-sq:                                           Obs per group:
     within  = 0.5191                                        min =        30
     between = 0.7177                                        avg =      30.0
     overall = 0.5661                                        max =        30

                                                F(6,429)           =     77.18
corr(u_i, Xb)  = -0.2950                        Prob > F           =    0.0000
```

zGR_Housepriceindex	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
zCPI						
L1.	-.2322964	.0372345	-6.24	0.000	-.3054812	-.1591116
zGR_GDP						
L1.	.1304268	.0367484	3.55	0.000	.0581975	.2026561
zGD_Population						
L2.	.3553218	.0858812	4.14	0.000	.1865215	.5241221
zGD_Unemploymentrate						
L2.	-.413835	.0481004	-8.60	0.000	-.5083768	-.3192931
zGD_Wages						
L6.	.1360996	.0347859	3.91	0.000	.0677277	.2044715
zInterestrate						
L4.	-.1467143	.0562664	-2.61	0.009	-.2573063	-.0361222
_cons	.0090732	.0286881	0.32	0.752	-.0473135	.0654599
sigma_u	.32182704					
sigma_e	.59755214					
rho	.22484451	(fraction of variance due to u_i)				

```
F test that all u_i=0: F(14, 429) = 6.50                     Prob > F = 0.0000
```

Annex 8 Hausman test of Model 6

```
. hausman fe re

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |      fe           re          Difference          S.E.
-------------+----------------------------------------------------------------
         CPI |
         L1. |   -1.037351    -.9879523        -.049399        .0418844
      GR_GDP |
         L1. |    .1594861     .1562317        .0032544        .0058637
GR_Populat~n |
         L2. |    3.286816     2.409228        .8775878        .5729606
GR_Unemplo~e |
         L2. |   -.1824018    -.1845703        .0021686        .0042115
    GR_Wages |
         L6. |    .1915771     .2033547       -.0117776        .0030431
Interestrate |
         L4. |   -.2913037      -.28412       -.0071837        .0497526
------------------------------------------------------------------------------
                           b = consistent under Ho and Ha; obtained from xtreg
            B = inconsistent under Ha, efficient under Ho; obtained from xtreg

    Test:  Ho:  difference in coefficients not systematic

                  chi2(6) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =        4.15
                Prob>chi2 =      0.6563
                (V_b-V_B is not positive definite)
```